Molecular and morphological clocks for estimating evolutionary divergence times

Jose Barba-Montoya^{1,2}, Qiqing Tao^{1,2}, and Sudhir Kumar^{1,2*}

¹*Institute for Genomics and Evolutionary Medicine, Temple University, Philadelphia, PA*

²*Department of Biology, Temple University, Philadelphia, PA*

*Corresponding author:

E-mail: s.kumar@temple.edu

Sudhir Kumar

Temple University

Philadelphia, PA 19122, USA

Abstract

Background: Matrices of morphological characters are frequently used for dating species divergence times in systematics. In some studies, morphological and molecular character data from living taxa are combined, whereas others use morphological characters from extinct taxa as well. We investigated whether morphological data produce time estimates that are concordant with molecular data. If true, it will justify the use of morphological characters alongside molecular data in divergence time inference.

Results: We systematically analyzed three empirical datasets from different species groups to test the concordance of dates of species divergence inferred using molecular and discrete morphological data from extant taxa as test cases. We found a high correlation between their divergence time estimates, despite a poor linear relationship between branch lengths for morphological and molecular data mapped onto the same phylogeny. This was because node-to-tip distances showed a much higher correlation than branch lengths, because of an averaging effect over multiple branches. We found that nodes with a large number of taxa often benefit from such averaging, but some considerable discordance between time estimates from molecules and morphology may still occur because some deeper branches show large difference from two types of data.

Conclusions: Our findings suggest that node- and tip-calibration approaches may be better suited for nodes with a large number of taxa. Nevertheless, we highlight the importance of evaluating the concordance of time structure in morphological and molecular data before any dating analysis using combined datasets.

Keywords: Species divergence time estimation, morphological clock, molecular clock, fossil calibration, Bayesian inference.

Background

There is a growing interest in using morphological data to infer species divergence times in systematics by using it itself in morphological clock analyses [e.g., 1–8] or by combining it with molecular data in total-evidence dating approaches [e.g., 9–14]. In some total-evidence dating, morphological characters from dated fossil species and extant species are analyzed along with molecular data under models of morphological character evolution [e.g., 15]. Tip-calibrations are typically used as constraints on the timing of lineage divergence by the inclusion of dated fossil species in the dating analysis [16] (Fig. 1). Moreover, in the fossilized birth-death (FBD) process [17, 18], the age of the fossil species provide the calibration information that translates the morphological distances into absolute times and rates, which are propagated to the other nodes on the phylogeny for which molecular data provides time structure information.

Nevertheless, utilizing morphological characters in a dating analysis is complicated [11, 16, 19–23]. To begin with, morphological traits are driven by natural selection and adaptation, they occasionally experience convergent evolution, and rarely evolve in a clock-like fashion [24, 25]. Furthermore, the sampling of morphological characters is generally biased towards taxonomically diagnostic traits. Therefore, morphological datasets consider mainly traits that are shared by multiple lineages, with a relatively small sampling of autapomorphies and invariant characters, which is known to have an impact on divergence time estimates [22]. Lastly, morphological datasets present extensive variation of evolutionary rates among branches in the phylogeny and generally relatively low numbers of characters sampled compared to phylogenomic datasets [21].

In this study, we investigated whether morphological data produce time estimates that are concordant with molecular data. It is essential for total-evidence studies in which morphological characters from extinct taxa are included, resulting in the morphological data being the only source of time information, for example in FBD dating approaches (Fig. 1). Comparisons of tip-calibration methods applying different FBD models/priors [e.g., 5, 11, 18, 19] have reported sharp conflict between molecular and morphological datasets under standard stochastic models causes total-evidence dating to produce older estimates when using inadequate models or vague FBD priors. Furthermore, the results are affected by different calibration approaches that lead to different divergence time estimates. The primary source of the problem remains unclear, which we wish to investigate by examining whether there is significant time structure in the morphological data.

We are interested in comparing the intrinsic time structure offered by morphological and molecular data with and without calibrations. The time structure will clearly be enhanced by the presence of calibrations, which has been the focus of tests by others [e.g., 5, 11, 18, 19]. Figure 1 illustrates how in fossil tip-calibration and FBD dating approaches, the phylogenetic position of a fossil (e.g., Fig. 1, species D*, E* or I*), and the time duration of the branch connecting it to the extant tree are determined based on morphological evidence alone in fossil tip-calibration and FBD

dating approaches. In order to reliably estimate t_4 and t_7 from Figure 1, the most basic requirement is that both morphology and molecules produce concordant time trees. If true, extrapolation involved in estimating t_4 and t_7 is possible because the molecular data are missing for extinct species. One way to evaluate this hypothesis is to assess the concordance of node-dates using the data from the living species. So, we analyzed three empirical datasets from different species groups, Hemiptera (true bugs), Hymenoptera (ants, bees, sawflies, wasps), and Spermatophyta (seed plants) to assess the relationship between morphological and molecular data. Moreover, we evaluated the impact of incorporating morphological characters into a joint analysis with molecular on divergence time estimates.

Figure 1. Integration of morphological characters from living and extinct species in a combined analysis with molecular data. In this case, the combined matrices contain morphological characters from both fossil (*) and living species, but no molecular data from fossil species. Since the branch lengths from fossils are only estimated by non-molecular characters, intuitively, we would expect them to be suitable for analysis along with living species only if the branch lengths from non-molecular data are proportional to molecular data.

Results

Molecular versus morphological maximum likelihood branch lengths

The maximum likelihood (ML) estimates of branch lengths for the same tree topology were obtained for molecular and morphological data separately. The three datasets – Hemiptera, Hymenoptera, and Spermatophyta – showed a limited correlation and high variation between the molecular and morphological branch lengths ($R^2 = 0.468, 0.252$ and 0.243 respectively; Figs. 2A, 2B, 2C). All three datasets showed rather short terminal morphology-based branch lengths as compared to the molecular data, including several terminal branches of zero length. Overall, the morphological trees differed from the molecular trees in having shorter terminal branch lengths. The morphological subsets analyzed might not have sampled autapomorphies as much as changes on internal branches, generating artefactually truncated terminal branches for the morphological tree [26].

Figure 2. Branch lengths and substitution model parameters were optimized on the same topology for both molecular and morphological data only for (A) Hemiptera, (B) Hymenoptera, and (C) Spermatophyta. The scatterplots to the right of the trees show the linear relationship between the branch lengths obtained from molecules versus morphology. The slope and coefficient of determination (R^2) for the linear regression through the origin are shown. The black dashed line represents the best-fit linear regression through the origin. The solid grey line represents equality between estimates.

Assessing molecular and morphological divergence time estimates

Analyses without internal calibrations

We first compared time estimates without applying internal calibrations (strategy Cr), which is critical to learn about the intrinsic time structure in the data. In the Hemiptera dataset, the correlation between the time estimates obtained from molecules and morphology was highly significant ($R^2 = 0.91$; Fig. 3A), which was in contrast to the pattern observed for branch lengths (Fig. 2A). Overall, 72% of morphological node times fell within the 95% credibility intervals (CIs) for molecular node times. Some of the nodes with significant differences are evident in Fig. 3A and Fig. S1A (nodes 48, 57, 58, 59, 60, 65, 67, 72, 73, 77 and 79). These nodes are connected to branches that showed extreme disagreement between morphological and molecular data.

In the Hymenoptera dataset the correlation between the time estimates obtained from molecules and morphology was also highly significant ($R^2 = 0.933$), but so was the variation (Fig. 3D).

Morphological data produced younger estimates than molecular, except for a few node times, which were just slightly older than molecular ages (Figs. 3D, S2A). Nodes that show a large difference (Figs. 3D and Fig. S2A; nodes 57, 58, 65, 68, 72, 73, 74, 75, 76, 77, 78, 84, 85 and 90) are related to highly disproportional branches between morphological and molecular data. This time, only 62% of morphological node times fell within the 95% CIs for molecular time estimates (Fig. S2A). In the Spermatophyta dataset the correlation between the time estimates obtained from molecules and morphology was again high ($R^2 = 0.937$), but the linear relationship was much tighter (Fig. 3G). This is interesting because the branch lengths showed very little correspondence (see Fig 2C). For most of the nodes in the timetree, morphological data produced younger estimates than molecular. The nodes with most significant differences (Figs. 3G and S3A; nodes 20, 21, 28, 29, 35) are related to highly discordant morphological and molecular branches. 88% of morphological node times fell within the 95% CIs for molecular time estimates (Fig. S3A).

The difference between the time estimates obtained from molecules and morphology is possibly because morphological matrices generally consist only of characters that vary among terminals; thus, the rate of morphological change, which ultimately impacts divergence time estimation, may be biased. Furthermore, it is generally accepted that phenotypic diagnostic changes of lineages will accumulate after their complete genetic isolation [27] while genetic changes may start well before the speciation time [8, 28, 29]. Therefore, the relative pace of the morphological clock should be slower than the molecular clock and thereby morphological time estimates could become younger than molecular time estimates.

Figure 3. The posterior mean times (empty black dots) and 95% CIs under calibration strategies Cr (green lines), C1 (red lines) and C2 (purple lines) for the molecular subsets are plotted against the morphological from Hemiptera, Hymenoptera and Spermatophyta datasets. The slope and coefficient of determination (R^2) for the linear regression through the origin are shown. The black dashed line represents the best-fit linear regression through the origin. The solid grey line represents equality between estimates.

Analyses with multiple calibrations

We expected the above results to improve when we use many calibrations. So, we used multiple calibration constraints but made them diffuse in our first analyses, reflecting agnosticism on the fossil ages, which is considered a more realistic scenario (strategy C1). Indeed, in the Hemiptera dataset the correlation between the time estimates obtained from molecules and morphology became higher under this calibration strategy ($R^2 = 0.966$; Fig. 3B). It is worth mentioning that, although differences were still found between age estimates from morphology and molecules, on average, timescales were similar (Fig. 3B, slope = 1.103). We found that 85% of morphological node times fell within the 95% CIs for molecular time estimates (Fig. S1B).

In the Hymenoptera dataset, the correlation between the time estimates obtained from molecules and morphology remained similar ($R^2 = 0.931$; Fig. 3E) than in strategy Cr. Morphological data produced younger estimates than molecular, except for a few node times, which were just slightly

older than molecular ages (Figs. 3E, S2B). 42% of morphological node times fell within the 95% CIs for molecular time estimates (Fig. S2B). In the Spermatophyta dataset the correlation between the time estimates obtained from molecules and morphology was very high ($R^2 = 0.984$; Fig. 3H). Morphological data produced nearly identical time estimates to molecular dates, except for node 28 that was very young compared to molecular dates (Figs. 3H, S3B). 94% of morphological node times fell within the 95% CIs for molecular time estimates (Fig. S3B).

Analyses with narrow calibrations

In many studies, however, investigators apply narrower calibration densities. So, we applied narrow calibration constraints in strategy C2, which reflects a prior assumption that fossil ages represent the “true” age of the corresponding lineage. We thus expected an even greater improvement of the above results when we use multiple narrow calibrations. In effect, in the Hemiptera dataset, the concordance of dates from morphological and molecular data became higher ($R^2 = 0.981$; Figs. 3C, S1C). The differences among molecular and morphological time estimates were the smallest for calibration strategy C2, showing a linear slope close to 1. Although differences were found between age estimates from morphology and molecules, on average, timescales were similar. Under strategy C1 85% of morphological node times fell within the 95% CIs for molecular time estimates, while this proportion increased to 96% under strategy C2 (Figs. 3B, 3C, S1).

In the Hymenoptera dataset, the concordance of dates from morphological and molecular data also improved ($R^2 = 0.96$; Figs. 3F, S2C). As under strategy C1, morphological data produced mostly younger estimates than molecular (with only a few exceptions; Fig. 3F). The differences among molecular and morphological time estimates were the smallest for calibration strategy C2, showing a linear slope close to 0.9, and 69% of morphological node times fell within the 95% CIs for molecular time estimates (Figs. 3F, S2C). In the Spermatophyta dataset the concordance of dates from morphological and molecular subsets remained high ($R^2 = 0.975$; Fig. 3I). As under strategy C1, morphological data produced nearly identical time estimates to molecular, except for node 28, that was very young compared to molecular dates (Figs. 3I, S3C). The differences among molecular and morphological time estimates were the smallest for calibration strategy C2, showing a linear slope close to 1 and 94% of morphological node times fell within the 95% CIs for molecular time estimates (Figs. 3I, S3C). The strong similarity between morphological and molecular time estimates under this calibration strategy with narrow probability densities occurs because nodes were constrained within a restricted period forcing the time prior and posterior estimates into agreement, so that both molecular and morphological data played a minor role in inferring divergence times [30–32].

Assessing divergence time estimates from combined datasets

Divergence time estimates from the Hemiptera combined subset were very similar to molecular estimates under the three calibration strategies, although morphological data produced slightly

younger estimates than molecular data under calibration strategies Cr and C1 (Figs. 4A, 5B). However, under both calibration strategies only a few node times deviated from a 1:1 linear trend (Figs. 4A, 5B). Under the three calibration strategies 100% of combined node times fell within the 95% CIs for molecular time estimates (Figs. 4A, 4B, 4C, S1). Time estimates from the Hymenoptera combined subset were almost identical to molecular estimates under the three calibration strategies, and only a few node times deviated from a 1:1 linear trend. Mainly, node 58 was consistently older for the molecular subset under the three strategies. A possible reason is that the ancestral and descendant branches of node 58 have extremely different lengths in the molecular and morphological trees (Fig. 2B), and this pattern persisted in the posterior time estimates. Under the three calibration strategies 100% of combined node times fell within the 95% CIs for molecular time estimates (Figs. 4D, 4DE, 4F, S2). Divergence time estimates from the Spermatophyta combined subset were nearly equal to molecular estimates under all calibration strategies, none of the nodes deviated from a 1:1 linear trend (Figs. 4G, 4H, 4I, S3). Overall, combining morphological and molecular datasets had little effect on divergence time estimates. The strong similarity between combined and molecular time estimates appears to be partly due the relatively small size of the morphological datasets and their low information content due to variable rates [16].

Figure 4. The posterior mean times (empty black dots) and 95% CIs under calibration strategies Cr (green lines), C1 (red lines) and C2 (purple lines) for the molecular subsets are plotted against the combined from Hemiptera, Hymenoptera and Spermatophyta datasets. The slope and coefficient of determination (R^2) for the linear regression through the origin are shown. The black dashed line represents the best-fit linear regression through the origin. The solid grey line represents equality between estimates.

Discussion

While some investigators have compared dates from morphological and molecular data, they have generally used multiple calibrations in their analyses [e.g., 1, 5, 8]. Overall, they found that both data sources resulted in comparable time estimates, except for some morphological estimates that were older than molecular, depending on the choice of calibration priors and data partitioning. Others have compared times from total evidence against node-dating [5, 10, 13, 33–35], and reported that in total-evidence divergence times are less sensitive to prior assumptions and tend to have smaller confidence/credibility intervals than in node-dating. The difference in time estimates between the two approaches could be related to differences in the priors used for node-dating and total-evidence dating, although there was a tendency under the total-evidence approach to producing older time estimates.

In this study, we explored the pattern of branch lengths from the morphological and molecular subsets based on ML estimates of branch lengths for the same topology for both molecular and morphological datasets (Figs. 3A, 3B, 3C). We found that morphological evolution along the tree was generally uncoupled from molecular evolution, as there was mostly low correlation and high dispersion between the molecular and morphological branch lengths in the three datasets analyzed (Hemiptera, Hymenoptera, and Spermatophyta). Morphological trees also differ from the molecular tree in having shorter terminal branch lengths, a pattern that was persistent in all three datasets. These issues appear to be partly because morphological characters may evolve at much more variable rates than molecules, but also it is very likely to be partly due to ascertainment bias. Thus, the disagreement between molecular and morphological branch lengths was expected to have a significant detrimental impact on the dating analysis.

However, the observed concordance between divergence times produced by these two types of data seems to be much more than that anticipated based on the comparison of individual branch lengths. The reason is that relative node-to-tip distances for morphological and molecular data show a much better relationship than the branch lengths (Fig. 5). The node-to-tip distance estimate for a node is the sum of the lengths of all paths from this node to all descendent tips divided by the total number of descendant tips. In the Hemiptera dataset the correlation between node-to-tip distances ($R^2 = 0.786$) was much higher than that for the branch lengths ($R^2 = 0.468$), although the relationship was still noisy (Fig. 5A). In the Hymenoptera and the Spermatophyta datasets the correlation between node-to-

tip distances was also rather high ($R^2 = 0.635$ and 0.591 respectively), and the dispersion was lower than in the comparison of branch lengths (Figs. 5B, 5C).

Figure 5. The linear relationship between node-to-tip distances from molecules alone versus morphology for (A) Hemiptera, (B) Hymenoptera, and (C) Spermatophyta. The black dashed line represents the best-fit linear regression through the origin. The slope and coefficient of determination (R^2) for the linear regression through the origin are shown. The black dashed line represents the best-fit linear regression through the origin. The solid grey line represents equality between estimates.

Therefore, a good relationship between node-to-tip distances from morphological and molecular data is the primary reason for the concordance between divergence times produced by these two types of data than that anticipated based on the comparison of individual branch lengths. The averaging of path lengths to multiple descendants in calculating node-to-tip distances is likely creating the time structure in both molecular and morphological data. If so, we would expect their relationship to be weaker for nodes with fewer descendants, as there are fewer evolutionary paths to average over. This effect was indeed observed in the Hemiptera dataset, terminal nodes showed much worse node-to-tip distance relationships, as they show significant deviation from the linear trend (Fig. 6). The terminal nodes represent the split of one lineage to form two living taxa for which the R^2 was only slightly higher in the node-to-tip comparison than that for the branch lengths.

The linear relationships of molecular and morphological node-to-tip distance estimates were higher for intermediate divergence nodes (Fig. 6), which were the nodes that had at least three descendants, but excluding the root and nodes that represented the deep divergence nodes set. Then, the linear relationships of molecular and morphological node-to-tip distance estimates of the deepest nodes were the highest – with an R^2 close to one. This suggests that for the Hemiptera dataset tip-calibration methods might be better suited for nodes with a large number of taxa. However, the linear relationships of molecular and morphological node-to-tip distance estimates of the deepest nodes were the lowest in the Hymenoptera and Spermatophyta datasets, indicating discordance of the relative deep branches. Therefore, for these two datasets tip-calibration methods might be better suited for intermediate and shallow nodes (Fig. 6).

Figure 6. Linear relationships of molecular and morphological estimates of node-to-tip distances for nodes that locate at the terminal (solid), intermediate (open) and deep (hatch) divergence regions for Hemiptera, Hymenoptera, and Spermatophyta. An R^2 of one indicates perfect linear relationship between molecular and morphological estimates, which is marked as the gray dashed line.

Because divergence times are a function of relative branch lengths [36, 37] rather than absolute values, time estimates were expected to become comparable under molecular and morphological clock models. In effect, the relationship between morphological and molecular divergence time estimates was much stronger, even when no internal calibrations were used (calibration strategy Cr). This suggests a much more concordant time structure in morphological and molecular datasets than that captured in the comparison of branch lengths. We hypothesized that this time structure would be reflected in the relative node-to-tip distances between morphological and molecular data, which ultimately dictate node ages [37].

Based on our results, it seems possible that duration of the branches connecting fossils to the extant tree, which are determined based on morphological evidence, may not be wrongly inferred for deepest and intermediate nodes in tip-calibration and FBD dating approaches (see also, by Püschel et al. (2019) [20]). Moreover, the high concordance of relative time estimates between morphological and molecular data when multiple calibrations are applied, and the extreme similarity of time estimates between combined and molecular data justifies the use of morphological data in node-, tip-calibration and FBD dating approach. Nevertheless, this kind of analysis need to carefully compare concordance of time estimates between datasets. As we showed discordance on time structure can only be corrected by calibrations and, still some nodes will remain different (e.g., Hemiptera nodes 58 and 59; Hymenoptera nodes 58 73, 74, 75, and 85; Spermatophyta node 28. Figs. SA-SC).

Conclusions

Overall, our study allows us to conclude that (1) although the relative ML branch lengths between morphological and molecular characters were very different, the relative node-to-tip distances were somewhat concordant suggesting a much more concordant time structure in the morphological and molecular dataset than that captured in the comparison of branch lengths. (2) When no internal calibrations were applied, morphological and molecular clock produced time estimates with a high correlation, although their absolute values were different. The concordance between time estimates was improved by applying multiple and/or narrow internal calibrations. (3) The combination of molecular and morphological data generally resulted in time estimates nearly identical to the ones from molecular data alone. Our study allows us to conclude that although there is concordant time structure in morphological and molecular data, the interpretation of fossil ages is determinant for the agreement of time estimates from morphological and molecular clock analyses. We emphasize the importance of evaluating the concordance of time structure in morphological and molecular data before any dating analysis using combined datasets.

Methods

Taxon sampling, molecular and morphological data, and tree topologies

For comparison between results from molecules and morphology, we composed three datasets consisting of extant species, for which both molecular and morphological data were available. Thus, extinct taxa were not included in our matrices. First, a dataset of 174 discrete morphological characters and 3731 molecular characters (base pairs — bp) from 44 Hemiptera species was obtained from Vea and Grimaldi (2016) [33]. Second, a dataset of 353 discrete morphological characters and 5096 molecular characters (bp) from 55 Hymenoptera species was taken from Ronquist et al. (2012) [10]. Third, a dataset of 121 discrete morphological characters and 19870 molecular characters (bp) from 18 Spermatophyta species was obtained from Doyle (2006) [38] and Morris et al. (2018) [39], respectively. Three rooted trees (Fig. 7) were constructed based on the literature: one for Hemiptera species [33]; one for Hymenoptera species [10]; and one for Spermatophyta species [39]. The three topologies from Figure 7 were fixed in all analyses to ensure consistent placement of calibrations and to avoid the confounding effects of alternative phylogenies and calibrations. Moreover, the use of a phylogeny reliably inferred as a fixed topology for dating is a common practice [e.g., 7, 37–40].

Figure 7. Phylogenies for (A) 44 Hemiptera species, (B) 56 Hymenoptera species, and (C) 18 Spermatophyta species. Calibration nodes are indicated by closed red dots.

Fossil calibrations and calibration strategies

Fossil calibrations are the foremost source of information for translating the distances between molecular/morphological sequences into estimates of absolute times and absolute rates in clock dating analysis. Thus, the quality of calibrations is expected to have a significant impact on divergence time estimates, even if a large amount of sequence data is available. Therefore, we employed three fossil calibration strategies simulating different quality of calibrations based on fossil evidence. Fossil calibrations were employed following justifications from the original studies, (Hemiptera [33], Hymenoptera [10], Spermatophyta [30]; Figure 7 and Table 1). For the three strategies, node calibrations were specified using a uniform distribution from fossil minimum-ages (Table 1).

Calibration strategy Cr. Only one calibration on the root was specified (no internal calibrations), assigning a uniform distribution borrowed from calibration strategy C1 as described below (Table 1). This strategy reflects a poor fossil record and the lack of fossil evidence for the internal nodes. However, it has the advantage of being the baseline for examining the impact of using calibration constraints on divergence time estimates.

Calibration strategy C1. Calibrations were specified using a uniform distribution $U(t_L, t_U)$, where t_L is the minimum age bound and t_U the maximum age bound, with mean $t_M = \frac{1}{2} (t_L + t_U)$. The offset and mean values from the original calibration were used respectively as the minimum bound (t_L) and mean (t_M) for the uniform distribution so that the maximum bound was $t_U = \frac{t_M}{\frac{1}{2}} - t_L$ (Table 1). This strategy reflects agnosticism about the true time of divergence between these bounds, which is considered a more realistic scenario.

Calibration strategy C2. Calibrations were also specified using a uniform distribution $U(t_L, t_U)$. The offset value from the original exponential distribution was used as the minimum bound (t_L),

and the maximum bound was specified at $t_U = t_L + \frac{t_L}{10}$, which assigns a probability of 10% for the minimum bound (t_L) to be older (Table 1). This strategy reflects a prior belief that the fossil minima are a close approximation of clade age. Such a strategy is only appropriate for calibration nodes within a restricted time, assuming that fossil ages represent the “true” age of the corresponding lineage.

Table 1. Summary of fossil calibrations and calibration strategies used in this study in million years before the present.

Dataset	Node	Clade	Minimum divergence time (Ma)	Calibration strategy Cr	Calibration strategy C1	Calibration strategy C2
Hemiptera	91 (root)	Coccoomorpha	140 († <i>Eomatsuccoccus casei</i>)	U(140,340)	U(140,340)	U(140,354)
Hemiptera	79	Kuwaniidae	45 († <i>Hoffeinsia foldi</i>)	—	U(45,155)	U(45,50)
Hemiptera	77	Ortheziidae	135 († <i>Cretoresia hamanaica</i>)	—	U(135,145)	U(135,149)
Hemiptera	74	Putoidae	45 († <i>Puto</i> sp.)	—	U(45,155)	U(45,50)
Hemiptera	72	Pseudococcidae	135 († <i>Williamsicoccus megalops</i>)	—	U(135,145)	U(135,149)
Hemiptera	64	Coccidae	98 († <i>Rosahendersonia prisca</i>)	—	U(98,122)	U(98,108)
Hemiptera	50	Diaspididae	50 († <i>Normarkicoccus cambayae</i>)	—	U(50,150)	U(50,55)
Hymenoptera	111 (root)	Hymenoptera	235 († <i>Triassoxyela, Asioxyela</i>)	U(235,369)	U(235,369)	U(235,259)
Hymenoptera	108	Tenthredinoidea	140 († <i>Palaeathalia</i>)	—	U(140,328)	U(140,154)
Hymenoptera	79	Vespina	180 († <i>Brigittepteris</i>)	—	U(180,290)	U(180,198)
Hymenoptera	78	Apocrita	176 († <i>Cleistogaster</i>)	—	U(176,294)	U(176,194)
Hymenoptera	69	Siricoidea	161 († <i>Aulisca</i>)	—	U(161,309)	U(161,177)
Hymenoptera	63	Pamphiloidea	161 († <i>Aulidontes, Pamphilidae</i> sp.)	—	U(161,309)	U(161,177)
Hymenoptera	57	Xylidae	180 († <i>Eoxyela</i>)	—	U(180,298)	U(180,198)
Spermatophyta	36 (root)	Spermatophytes	308.14 († <i>Cordaites iowensis</i>)	U(308,366)	U(308,366)	U(308,339)
Spermatophyta	35	Angiosperms	125.9 (tricolpate pollen)	—	U(126,247)	U(126,139)
Spermatophyta	31	Stem-Saururus	44.3 († <i>Saururus tuckerae</i>)	—	U(44,247)	U(44,48)
Spermatophyta	30	Acrogymnosperms	308.14 († <i>Cordaites iowensis</i>)	—	U(308,366)	U(308,339)
Spermatophyta	29	Gingko-Cycas	264.7 († <i>Crossozamia</i>)	—	U(265,366)	U(265,282)
Spermatophyta	27	Conifers	147 († <i>Rissikia media</i>)	—	U(147,312)	U(147,162)
Spermatophyta	20	Gnetales	119.6 († <i>Eoantha zherkihini</i>)	—	U(120,312)	U(120,132)

Nodes are numbered as in Figure 7. Fossil taxa are indicated by a dagger (†) before their names.

Modeling morphological and molecular evolution

To model morphological evolution we used the Mk model [15], with the ascertainment bias set to variable (only variable characters scored), equal state frequencies and assuming discrete gamma-distributed heterogeneity among sites (Mk + Γ_5). To model molecular evolution, we used the GTR general time-reversible (GTR) model [43], with discrete gamma-distributed heterogeneity among sites (GTR + Γ_5). Unlike the model applied for morphological data (Mk + Γ_5), the GTR + Γ_5 model relaxes the assumption of equal character frequencies in molecular substitutions. A ML optimization of model parameters and branch lengths for the given topology was performed for both molecular and morphological data separately in MEGAX [44] and RAXML [45] respectively, for the Hemiptera, Hymenoptera, and Spermatophyta datasets to assess the pattern of branches lengths from the morphological and molecular subsets. We also calculated the node-to-tip distances using the ML trees. For each node, the node-to-tip distance is the sum of the lengths of all paths from this node to all descendent tips divided by the total number of descendant tips.

Bayesian divergence time estimation

In each data analysis (Hemiptera, Hymenoptera, and Spermatophyta), we used one molecular subset, one morphological subset, and one combined (morphological + molecular) subset, and applied three calibration strategies, Cr, C1, and C2. All Bayesian analyses (27 in total) were carried out with the program MrBayes [46]. The sequence likelihood of the morphological and molecular subsets was calculated under the Mk + Γ_5 and the GTR+ Γ_5 substitution models, respectively. The combined subset was treated as two partitions under separate substitution models (Mk + Γ_5 , and GTR+ Γ_5), with separate substitution-rate parameters assigned and estimated for each partition. The MCMC sampling settings were determined through pilot runs and differed among the datasets. We ran each analysis at least twice and checked for convergence by comparing the posterior mean estimates between runs and by plotting the time series traces of the samples. We then merged the samples from the runs before summarizing the posterior. Moreover, we ran analyses without sequence data to obtain an estimate of the effective time prior that informed the resulting marginal priors for the node ages after truncation. We plotted these alongside node age estimates from the posteriors to assess how much the data is influencing age estimates (Figs. S1, S2, S3). The prior settings and the MCMC run parameters of each dataset analysis are detailed below.

Analysis of Hemiptera dataset

Analyses were performed on three subsets from 44 Hemiptera species, one of 174 morphological characters with 12.15% of missing data; one of 3731 molecular characters (bp) with 47.96% of missing data; and one combined subset (morphological + molecular) of 3905 characters with 46.36% of missing data. The rooted tree topology with seven calibration nodes (Fig. 7A) was fixed in all analyses. Detailed information on the calibrations is given in Table 1. Models of evolution for each subset were implemented, as described earlier. Priors for rates were set as in Vea and Grimaldi (2016) [33]. The *Independent Gamma Rate* (IGR) model, in which the rates of branches varied independently from a gamma distribution [47] was used as a rate prior. The gamma model is parametrized using two parameters: the mean and variance. The time unit was 100 Myr. The mean was assigned a lognormal hyperprior LN(-6.0605,0.0519), with the mean of $\exp\{-6.0605 + 0.0519^2/2\} = 0.0023$, and the variance (*Igrvarpr*) was assigned an exponential hyperprior with mean 0.039. For each analysis four MCMC runs were performed, each consisting of 5×10^6 iterations, sampling every 200, with the burn-in set to 10%, resulting in a total of 9×10^4 samples from the four runs, which were used to obtain posterior time estimates.

Analysis of Hymenoptera dataset

Analyses were performed on three subsets from 56 Hymenoptera species; one of 353 morphological characters with 19.97% of missing data; one of 5096 molecular characters (bp) with 25.65% of missing data; and one combined subset (morphological + molecular) of 5449 characters with 25.28%

of missing data. The rooted tree topology with seven calibration nodes (Fig. 7B) was fixed in all analyses. Detailed information on the calibrations is given in Table 1. Models of evolution for each subset were implemented, as described earlier. Priors for rates were set as in Ronquist et al. (2012) [10] for the three subsets. The IGR model was used to specify prior for rates. The time unit was 100 Myr. The mean of the gamma was assigned a lognormal hyperprior LN (-7.1,0.5), with the mean $\exp\{-7.1 + 0.5^2/2\} = 0.001$, and the variance of the gamma was assigned an exponential hyperprior with mean 0.027. Four MCMC runs were performed, each consisting of 5×10^6 iterations, sampling every 200, with the burn-in set to 10%, resulting in a total of 9×10^4 samples from the four runs.

Analysis of Spermatophyta dataset

Analyses were performed on three subsets from 18 Spermatophyta species; one of 121 morphological characters with 27.5% of missing data; one of 19870 molecular characters (bp) with 14.11% of missing data; and one combined subset (morphological + molecular) of 5449 characters with 14.18% of missing data. The rooted tree topology with seven calibration nodes (Fig. 7C) was fixed in all analyses. Detailed information on the calibrations is given in Table 1. Models of evolution for each subset were implemented, as described earlier. Priors for rates were set as in Barba-Montoya et al. (2018) [30] for the three subsets. The IGR model was used to specify prior for rates. The time unit was 100 Myr. The mean of the gamma was assigned a lognormal hyperprior LN (-2.79, 0.5), with the mean $\exp\{2.79 + 0.5^2/2\} = 0.07$, and the variance of the gamma is assigned an exponential hyperprior with mean 0.1. Six MCMC runs were performed, each consisting of 1×10^7 iterations, sampling every 200, with the burn-in set to 10%, resulting in a total of 2.7×10^5 samples from all six runs.

Abbreviations

FBD: fossilized birth-death; ML: maximum likelihood; CIs: credibility intervals; bp: base pairs; GTR: general time-reversible.

Supplementary information

Supplementary information accompanies this paper at <https://doi.org/...>

Declarations

Acknowledgments

We thank Aishwarya Durgam for technical support and Dr. Beatriz Mello for comments and editorial suggestions.

Author's contributions

SK and JBM designed the study. JBM and QT assembled the datasets. JBM performed the analyses. JBM, QT and SK wrote the manuscript. All authors edited the manuscript and approved the final version.

Funding

This research was supported in part by grants from the National Institutes of Health (NIH GM0126567-03), National Science Foundation (NSF 1661218), and National Aeronautics and Space Administration (NASA NNX16AJ30G) to S.K.

Availability of data and materials

The datasets generated and/or analyzed during the current study are available in the Figshare repository, <https://figshare.com...>

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Institute for Genomics and Evolutionary Medicine, Temple University, Philadelphia, PA.

²Department of Biology, Temple University, Philadelphia, PA

References

1. Schrago CG, Mello B, Soares AER. Combining fossil and molecular data to date the diversification of New World Primates. *J Evol Biol.* 2013;26:2438–46.
2. Beck RMD, Lee MSY. Ancient dates or accelerated rates? Morphological clocks and the antiquity of placental mammals. *Proc R Soc B Biol Sci.* 2014;281.
3. Lee MSY, Cau A, Naish D, Dyke GJ. Morphological clocks in paleontology, and a mid-cretaceous origin of crown aves. *Syst Biol.* 2014;63:442–9.
4. Matzke NJ, Wright A. Inferring node dates from tip dates in fossil Canidae: The importance of tree priors. *Biol Lett.* 2016;12.
5. Puttick MN, Thomas GH, Benton MJ. Dating placentalia: Morphological clocks fail to close the molecular fossil gap. *Evolution.* 2016;70:873–86.
6. King B, Qiao T, Lee MSY, Zhu M, Long JA. Bayesian Morphological Clock Methods Resurrect Placoderm Monophyly and Reveal Rapid Early Evolution in Jawed Vertebrates. *Syst Biol.* 2017;66:499–516.
7. Álvarez-Carretero S, Goswami A, Yang Z, Dos Reis M. Bayesian Estimation of Species Divergence Times Using Correlated Quantitative Characters. *Syst Biol.* 2019;68:967–86.
8. Caldas I V., Schrago CG. Data partitioning and correction for ascertainment bias reduce the uncertainty of placental mammal divergence times inferred from the morphological clock. *Ecol Evol.* 2019;9:2255–62.
9. Pyron RA. Divergence time estimation using fossils as terminal taxa and the origins of lissamphibia. *Syst Biol.* 2011;60:466–81.
10. Ronquist F, Klopstein S, Vilhelmsen L, Schulmeister S, Murray DL, Rasnitsyn AP. A total-evidence approach to dating with fossils, applied to the early radiation of the hymenoptera. *Syst Biol.* 2012;61:973–99.
11. Ronquist F, Lartillot N, Phillips MJ, Lyon CB. Closing the gap between rocks and clocks using total-evidence dating. *Philos Trans R Soc London B.* 2016;371:20150136.
12. Wood HM, Matzke NJ, Gillespie RG, Griswold CE. Treating fossils as terminal taxa in divergence time estimation reveals ancient vicariance patterns in the palpimanoid spiders. *Syst Biol.* 2013;62:264–84.
13. O'Reilly JE, dos Reis M, Donoghue PCJ. Dating Tips for Divergence-Time Estimation. *Trends Genet.* 2015;31:637–50.
14. Gavryushkina A, Heath TA, Ksepka DT, Stadler T, Welch D, Drummond AJ. Bayesian total-evidence dating reveals the recent crown radiation of penguins. *Syst Biol.* 2017;66:57–73.
15. Lewis PO. A likelihood approach to estimating phylogeny from discrete morphological character data. *Syst Biol.* 2001;50:913–25.
16. dos Reis M, Donoghue PCJ, Yang Z. Bayesian molecular clock dating of species divergences in the genomics era. *Nat Rev Genet.* 2016;17:71–80. doi:10.1038/nrg.2015.8.
17. Heath T a, Huelsenbeck JP, Stadler T. The fossilized birth–death process for coherent calibration of divergence-time estimates. *Proc Natl Acad Sci.* 2014;111:E2957–66. doi:10.1073/pnas.1319091111.
18. Zhang C, Stadler T, Klopstein S, Heath TA, Ronquist F. Total-Evidence Dating under the Fossilized Birth-Death Process. *Syst Biol.* 2016;65:228–49.
19. Luo A, Duchene D, Zhang C, Zhu C-D, Ho SYW. A Simulation-Based Evaluation of Total-Evidence Dating Under the Fossilized Birth-Death Process. *bioRxiv.* 2018;:436303. doi:10.1101/436303.

20. Püschel HP, O'Reilly JE, Pisani D, Donoghue PCJ. The impact of fossil stratigraphic ranges on tip-calibration, and the accuracy and precision of divergence time estimates. *Palaeontology*. 2019;:1–17.
21. Klopstein S, Ryer R, Coiro M, Spasojevic T. Mismatch of the morphology model is mostly unproblematic in total-evidence dating: insights from an extensive simulation study Seraina. *bioRxiv*. 2019.
22. Bromham L, Duchêne S, Hua X, Ritchie AM, Duchêne DA, Ho SYW. Bayesian molecular dating: opening up the black box. *Biol Rev*. 2018;93:1165–91.
23. Goloboff PA, Pittman M, Pol D, Xu X. Morphological data sets fit a common mechanism much more poorly than DNA sequences and call into question the Mk model. *Syst Biol*. 2019;68:494–504.
24. Barba-Montoya J, dos Reis M, Yang Z. Comparison of different strategies for using fossil calibrations to generate the time prior in Bayesian molecular clock dating. *Mol Phylogenet Evol*. 2017;114.
25. Kimura M. The neutral theory of molecular evolution. In: *Molecular Evolutionary Rates Contrasted With Phenotypic Evolutionary Rates*. Cambridge University Press; 1983. p. 55–97.
26. Lee MSY. Multiple morphological clocks and total-evidence tip-dating in mammals. *Biol Lett*. 2016;12.
27. Nei M, Kumar S. *Molecular Evolution and Phylogenetics*. Oxford: Oxford University Press.; 2000.
28. Edwards S, Beerli P. Perspective: Gene divergence, population divergence, and the variance in coalescence time in phylogeographic studies. *Evolution*. 2000;54:1839–54.
29. Capellão RT, Costa-Paiva EM, Schrago CG. Appropriate assignment of fossil calibration information minimizes the difference between phylogenetic and pedigree mutation rates in humans. *Life*. 2018;8:1–11.
30. Barba-Montoya J, dos Reis M, Schneider H, Donoghue PCJ, Yang Z. Constraining uncertainty in the timescale of angiosperm evolution and the veracity of a Cretaceous Terrestrial Revolution. *New Phytol*. 2018;218:819–34.
31. Magallón S, Gómez-Acevedo S, Sánchez-Reyes LL, Hernández-Hernández T. A metacalibrated time-tree documents the early rise of flowering plant phylogenetic diversity. *New Phytol*. 2015;207:437–53.
32. Magallón S, Hilu KW, Quandt D. Land plant evolutionary timeline: Gene effects are secondary to fossil constraints in relaxed clock estimation of age and substitution rates. *Am J Bot*. 2013;100:556–73.
33. Veá IM, Grimaldi DA. Putting scales into evolutionary time : the divergence of major scale insect lineages (Hemiptera) predates the radiation of modern angiosperm hosts. *Sci Rep*. 2016;6:1–11. doi:10.1038/srep23487.
34. Cascini M, Mitchell KJ, Cooper A, Phillips MJ. Reconstructing the evolution of giant extinct kangaroos: Comparing the utility of DNA, morphology, and total evidence. *Syst Biol*. 2019;68:520–37.
35. Arcila D, Pyron RA, Tyler JC, Ortí G, Betancur R. An evaluation of fossil tip-dating versus node-age calibrations in tetraodontiform fishes (Teleostei: Percomorphaceae). *Mol Phylogenet Evol*. 2015;82:131–45. doi:10.1016/j.ympev.2014.10.011.
36. Tamura K, Battistuzzi FU, Billings-Ross P, Murillo O, Filipowski A, Kumar S. Estimating divergence times in large molecular phylogenies. *Proc Natl Acad Sci U S A*. 2012;109:19333–8.
37. Tamura K, Tao Q, Kumar S. Theoretical foundation of the reltime method for estimating divergence times from variable evolutionary rates. *Mol Biol Evol*. 2018;35:1770–82.

38. Doyle JA. Seed ferns and the origin of angiosperms. *J Torrey Bot Soc.* 2006;133:169–209.
39. Morris JL, Puttick MN, Clark JW, Edwards D, Kenrick P, Pressel S, et al. The timescale of early land plant evolution. *Proc Natl Acad Sci.* 2018;:201719588.
40. Reis M Dos, Gunnell GF, Barba-Montoya J, Wilkins A, Yang Z, Yoder AD. Using Phylogenomic Data to Explore the Effects of Relaxed Clocks and Calibration Strategies on Divergence Time Estimation: Primates as a Test Case. *Syst Biol.* 2018;0 February:1–22. doi:10.1093/sysbio/syy001.
41. Oliveros CH, Field DJ, Ksepka DT, Keith Barker F, Aleixo A, Andersen MJ, et al. Earth history and the passerine superradiation. *Proc Natl Acad Sci U S A.* 2019;116:7916–25.
42. Li HT, Yi TS, Gao LM, Ma PF, Zhang T, Yang JB, et al. Origin of angiosperms and the puzzle of the Jurassic gap. *Nat Plants.* 2019;5:461–70. doi:10.1038/s41477-019-0421-0.
43. Tavaré S. Some probabilistic and statistical problems on the analysis of DNA sequences. *Lect Math life Sci.* 1986;17:57–86.
44. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol.* 2018;35:1547–9.
45. Stamatakis A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics.* 2014;30:1312–3.
46. Ronquist F, Teslenko M, Van Der Mark P, Ayres DL, Darling A, Höhna S, et al. Mrbayes 3.2: Efficient bayesian phylogenetic inference and model choice across a large model space. *Syst Biol.* 2012;61:539–42.
47. Lepage T, Bryant D, Philippe H, Lartillot N. A general comparison of relaxed molecular clock models. *Mol Biol Evol.* 2007;24:2669–80.

Figure S1. Calibration densities (dark grey bands), 95% CIs in the time prior (light grey bands), and posterior (colored lines) for 47 nodes in the Hemiptera timetrees under calibration strategies (A) Cr-green, (B) C1-red and (C) C2-purple. Calibrated nodes are numbered as in Figure 7.

617

618 **Figure S2.** Calibration densities (dark grey bands), 95% CIs in the time prior (light grey bands), and

619 posterior (colored lines) for 55 nodes in the Hymenoptera timetrees under calibration strategies (A)

620 Cr-green, (B) C1-red and (C) C2-purple. Calibrated nodes are numbered as in Figure 7.

621

623

624

625

626

Figure S3. Calibration densities (dark grey bands), 95% CIs in the time prior (light grey bands), and posterior (colored lines) for 17 nodes in the Spermatophyta timetrees under calibration strategies (A) Cr-green, (B) C1-red and (C) C2-purple. Calibrated nodes are numbered as in Figure 7.